

Curriculum Vitae

Paolo Manghi

March 2, 2012

Date of birth: 22nd of December, 1970

Place of birth: Firenze

Status: Unmarried

Children: 1

Nationality: Italian

Contacts: *Address:* Via Y. Gagarin, 85 - 56023 Cascina (PI), Italia

Office Tel.: +39 050 3152038

Office Fax: +39 050 3153464

E-mail: paolo.manghi@isti.cnr.it

Web: <http://www.isti.cnr.it/People/P.Manghi>

1 Introduction

Paolo Manghi is a Research Staff at Istituto di Scienza e Tecnologie dell'Informazione (ISTI) of Consiglio Nazionale delle Ricerche (CNR), Pisa, Italy. He is a member of the Digital Library (D-Lib) research group (led by Dr. Donatella Castelli), part of the Networked Multimedia Information Systems (NeMIS) laboratory (led by Dr. Fausto Rabitti).

Studies: He obtained a four-year doctorate in Informatica (Università di Pisa) and two MSC degrees: Laurea in Scienze dell'Informazione (Università di Pisa) and Laurea Specialistica in Tecnologie Informatiche (Università di Pisa).

Research: His curricula witnesses a research path that started from *programming and typing in “orthogonally” persistent languages*, to apply the same concepts on the Web/Internet, investigate *type correctness of query languages for XML databases* and *type-safe integration of programming languages and XML languages*, and explored *peer-to-peer service architectures for distributed XML databases*. From these experiences on data management and distributed service architectures he approached his today main interests: (i) *Digital Library Management Systems* for the realization of multimedia and repository archives, and (ii) *enabling services, functionality services and paradigms for service-oriented data infrastructures*.

Roles: He is *Research and Technical Manager* for the EU projects OpenAIRE [Pr1], OpenAIREplus [Pr3], and EFG1914 [Pr4]. He is appointed as *Project Technical Coordinator (WP5)* of the EU project HOPE [Pr2]. Details on the mentioned projects will follow.

Activities: He was and is also extensively involved in international cooperations and initiatives, authoring and hearings of several EC project proposals, project research and technical coordination, organization and editorship of international workshops, academic and professional teaching, supervision of university and post-doc students.

2 Studies

- 19.07.1996** Laurea (Degree MSc) in *Scienze dell'Informazione*, University of Pisa, final mark 110/110. Master Thesis: *Aspetti Linguistici della Costruzione di Applicazioni Persistenti*; supervisor: Prof. Giorgio Ghelli.
- 26.10.2001** Four-year Doctorate in *Informatica*, Pisa Consortium, XIIth Series. PhD Thesis: *Extracting Typed Values from Semistructured Databases*; supervisor: Prof. Giorgio Ghelli; referees: Prof. M. Atkinson (Glasgow, UK), Prof. F. Crestani (Glasgow, UK).
- 12.12.2003** Laurea (Degree MSc) *Specialistica in Tecnologie Informatiche*, University of Pisa, final mark 110/110.

3 Positions

Paolo Manghi started his research experience in 1996, in the academia, with a four-year doctorate between the University of Pisa and the Universities of Glasgow and Strathclyde in Scotland (UK) partly funded by a European Marie Curie TMR EC Research Training Fellowship grant. After three years in Scotland, from 1997 to 2000, working with Prof. Richard Connor, he returned to the University of Pisa where he re-joined the Fibonacci research group of Prof. Giorgio Ghelli and Prof. Antonio Albano until May 2006. Since May 2006, he moved at ISTI-CNR, to work with the D-Lib research group led by Dr. Donatella Castelli. Currently, he has a temporary position (art.23) as a researcher at CNR and has recently obtained the qualification (so-called “idoneità”) for a permanent researcher position.

3.1 Experience in Academia: 1995 - 2005

- September 1996 - University of Pisa** PhD Student Grant (01.09.1996-31.10.1997) at the Computer Science Department of the University of Pisa (reference: Prof. G. Ghelli).
- October 1997 - University of Glasgow** Research Fellow position (31.10.1997-30.06.1999), as European TMR Marie Curie Grant Holder, at the Department of Computer Science of the University of Glasgow (UK). Funding project: *Hippo* [Pr10], led by Prof. R. Connor, in cooperation with the Dipartimento di Informatica of Pisa University (reference: Prof. G. Ghelli).
- July 1999 - University of Strathclyde** Research Fellow position (01.07.1999-31.10.2000), as European TMR Marie Curie Grant Holder, at the Department of Computer and Information Sciences of Strathclyde University (UK). Project *SNAQue* [Pr11], led by Prof. R. Connor, in cooperation with the Dipartimento di Informatica of Pisa University (reference: Prof. G. Ghelli).
- November 2000 - University of Pisa** Research contract (29.11.2000-30.05.2001) for the research project *DataX*, coordinated by Prof. G. Ghelli, Dipartimento di Informatica, University of Pisa.
- June 2001 - University of Pisa** Research Fellow position (11.06.2001-11.06.2003), as *assegnista di ricerca*, at Dipartimento di Informatica, University of Pisa. Funding project: *Design and implementation of a query language for semistructured-data*, led by Prof. G. Ghelli.
- July 2003 - University of Pisa** Research contract (01.07.2003 - 30.09.2003) for the project *FIRB: Enabling Platforms for High-performance Computational Grids Oriented to Scalable Virtual Organizations (grid.it)* on the activity: *design of a distributed XML database over the GRID*. Project coordinator: Prof. M. Vanneschi, Dipartimento di Informatica, University of Pisa.
- October 2003 - Ministry of Education** Consulting for CRUI Foundation, Ministry of Education (01.10.2003-01.05.2005). Funding project: IT4PS (*Information Technology for Problem Solving*). Definition of lectures, final tests, and drafting of studying books for ECDL (European Computer Driving License) advanced university courses on MS Excel and MS Access, specific to Pharmacy and Medical Faculties.
- November 2003 - University of Pisa** Research Fellow position (01.11.2003-31.10.2005), as *assegnista di ricerca*, at Dipartimento di Informatica, University of Pisa, government funds under INF/01.

3.2 Experience at Consiglio Nazionale delle Ricerche: 2006 - ongoing

February 2006 - Consiglio Nazionale delle Ricerche Research contract (22.02.2006-22.03.2006) for the project *BELIEF* [Pr5] (led by Donatella Castelli) on the activity: *Definition of a document model for the representation of heterogeneous documents stored into European Institutional Repositories; design of a virtual collection service for such documents*, Project reference: Dr. Donatella Castelli, ISTI, Consiglio Nazionale delle Ricerche.

May 2006 - Consiglio Nazionale delle Ricerche Research Fellow position (01.05.2006-30.06.2008), as *assegnista di ricerca*, at ISTI, Consiglio Nazionale delle Ricerche, Pisa. Funding EC project: BELIEF [Pr5].

July 2008 - Consiglio Nazionale delle Ricerche Research Fellow position (01.07.2008-30.06.2009), as *ricercatore a tempo determinato* (art.23), at ISTI Consiglio Nazionale delle Ricerche, Pisa. Funding EC Project: DRIVER-II [Pr7].

March 2009 - Consiglio Nazionale delle Ricerche Qualification (*idoneità*) for a permanent researcher position (31.03.2009). Concorso da ricercatore, Procedura Comparativa R.07.01 - INF/01 (CNR).

July 2009 - Consiglio Nazionale delle Ricerche Research Fellow position (01.07.2009-30.06.2010), as *ricercatore a tempo determinato* (art. 23), at ISTI Consiglio Nazionale delle Ricerche, Pisa. Funding EC Project: DRIVER-II [Pr7].

July 2010 - Consiglio Nazionale delle Ricerche Research Fellow position (01.07.2010-31.12.2010), as *ricercatore a tempo determinato* (art. 23), at ISTI Consiglio Nazionale delle Ricerche, Pisa. Funding EC Project: OpenAIRE [Pr1].

January 2011 - Consiglio Nazionale delle Ricerche Research Fellow position (01.01.2011-30.06.2011), as *ricercatore a tempo determinato* (art. 23), at ISTI Consiglio Nazionale delle Ricerche, Pisa. Funding EC Project: OpenAIRE [Pr1].

July 2011 - Consiglio Nazionale delle Ricerche Research Fellow position (01.07.2011-15.09.2011), as *ricercatore a tempo determinato* (art. 23), at ISTI Consiglio Nazionale delle Ricerche, Pisa. Funding EC Project: OpenAIRE [Pr1].

September 2011 - Consiglio Nazionale delle Ricerche Research Staff position (16.09.2011-ongoing), as *ricercatore a tempo indeterminato III livello*, at ISTI Consiglio Nazionale delle Ricerche, Pisa.

4 Research

Paolo Manghi research activities gravitate towards the areas of database systems, query and programming languages, type systems, service architectures and infrastructures, digital library systems, with particular focus on:

- Programming and typing in “orthogonally” persistent languages;
- Type-safe integration of programming languages and XML languages;
- Type correctness of query languages for XML database;
- Peer-to-peer service architectures for distributed XML databases;
- Data Models and Types for Digital Library Management Systems;
- Data curation in Digital Libraries;
- Service-oriented (ICT) data infrastructures;
- Architectures for collaborative research environments.

This section describes these research topics and lists the research projects that were directly or indirectly related with them.

4.1 Topics

1995 - 1998 *Programming and Typing in “orthogonally” persistent languages.*

Orthogonally persistent (OP) languages provide run-time support for persistent storage of values, independently of their type. The idea is to blur the separation between run-time and persistent values, in order to avoid the extensive coding required to store and retrieve data on disk (estimated as 20%). From 1995 until 1996 he investigated the problems arising applying module programming techniques in the design of very large applications with OP languages. In OP languages, values, including module implementations, can be type-safely linked and modified at run-time into a computation. In traditional module-structured applications, changes to implementations usually require costly operations, such as recompilation and relinking. The study led to the definition of specific application construction methodologies that support module-systems separate compilation and permit, at the same time, run-time modifications to module implementations, so to limit application re-construction operations.

Along the same line, from 1996 until 1998, in cooperation with Prof. Richard Connor’s research group of Glasgow University (UK), he studied the possibility of integrating the notion of orthogonal persistence within the context of the Web. Indeed, the Web offers a series of standard languages and protocols, such as HTML, XML, and HTTP, which combined enable a direct way to store and share permanent data over distributed file systems. *Hippo*, High-level Internet Programming with Persistent Objects (EPSRC project, Advanced Fellowship GR/K 79222 - High Level Internet Programming Systems), is a typed procedural language whose applications store persistent data, together with their types, into local ad-hoc HTML documents uniquely identified over a persistent namespace over HTTP. Such data can be accessed and retrieved in order to be visualized within a browser or type-safely imported into a remote Hippo computation as a persistent typed value. The project extended the concept of orthogonal persistence to the Web, then naturally moved towards novel Internet programming paradigms, having non-determinism and autonomy as main research avenues.

Publications: [Man96] [CGM96] [CMS98]

1998 - 2001 *Type-safe integration of programming languages and XML languages.*

XML data collections are often generated for data publication or sharing within the context of traditional mainstream language computations or database applications. Accordingly, when such data reaches the target computation, one would be able to operate over it by means of the same tools and frameworks. On the contrary, XML data are mainly manipulated as labeled trees, within the declarative algebras of XML query languages or within the procedural algebras of mainstream languages APIs, such as DOM. In both cases: (i) the sender computation generates a value with a structure (type) that well fits the information to be modeled and enables static type checking; (ii) the receiver operates over the labeled tree representation of such value, where trees are not always the best structure to model the information source, and loses the type checking quality provided at the sender, which locally applies to the labeled tree structure.

Since 1998 until 2001 he worked, in cooperation with Fabio Simeoni and Prof. Richard Connor of Strathclyde University, on the *SNAQue* project [Pr11]. SNAQue project’s aim is the theoretical design and implementation of a system API (for Java language, but any language can implement it) for extracting typed values from XML data and import them type-safely within a run-time computation. Programmers can therefore operate over XML transmitted data, by fully recovering the benefits of complex type structures, in terms of modeling and type safety. The resulting framework inspired the definition of novel kinds of programming language run-time environments, enabling the realization of “hybrid” applications. In such environments, values can be manipulated as language values or as XML values of equivalent semantics, in a type-safe manner, by means of language constructs or XQuery constructs, according to the specific programming needs.

Publications: [Man01] [SML⁺02] [SLCM03] [MSLC02] [CLM⁺01]

1998 - 2004 *Type-Correctness of query languages for XML database.*

By definition, XML data may not be described by a schema. Accordingly, the run-time semantics of XML query languages is typically defined in a greedy way: a query always returns a result, even if its structure cannot be satisfied by the input data; in such case the query evaluates to an empty value. For the same reason, query correctness and type checking algorithms for XML query languages were considered irrelevant research issues. This was in a sharp departure from traditional query languages,

where a structural error generates a run-time error, query correctness is defined in terms of run-time errors avoidance, and type checking algorithms notify the errors to the programmers.

Lately, XML has been increasingly used in database systems and programming languages as the standard language for data publication and sharing. Such data comes with a schema, thus programmers may expect to define queries fed with input featuring the same structure. As a consequence, there has been a growing interest in techniques to identify parts of the query that will never contribute to the construction of the result.

Since 2000, he worked on the definition of a characterization of query correctness for XML query languages, with the aim of realizing a type system capable to identify at static time all possible structural anomalies. A first prototype of the type system has been designed for the language *TeQuyLa*, targeting search within XML representations of historical books and documents (e.g. Divina Commedia). Later on, a finer definition of correctness, namely *FE-correctness*, has been given for the query language μXQ [Pr14], designed to be the FLWR core language of XQuery (XML query language defined by the W3C). μXQ 's type system is proved sound and (partially) complete with respect to the formal definition of FE-correctness. In parallel, similar studies on types and expressiveness of queries have been conducted in the context of *TQL* [Pr13], an ambient calculus-based query language for semi-structured data.

Publications: [ADG⁺01] [CGMS06a] [ACG⁺00] [CGMS02] [CGA⁺02a] [CGMS04a] [CGA⁺02b]

2003 - 2007 *Peer-to-peer service architectures for distributed XML databases.*

Much of the research on the construction of real peer-to-peer (P2P) databases is currently aimed at the P2P decentralization of data integration mediators (e.g. Piazza and CoDB). These systems, usually based on the GLAV paradigm, enable each peer to reformulate queries according to a given set of mappings, with no need of centralized mediator. However, the need to set up the translation implies that some human administrative work is needed in order to join such systems. Furthermore, for such an effort to be worthy, such systems are not apt for highly volatile peers.

Since 2003 he worked on the design and implementation of *XPeer* [Pr15], an XML peer-to-peer (P2P) database system. XPeer differs from standard P2P database systems for two main key ideas. First, XPeer decouples query compilation, demanded to an *overlay network* of administrative nodes, from query execution, which is directly managed by peers. Compilation and execution decoupling allows for a better handling of *complex FLWR queries* (e.g. nested queries), and allows the overlay network to perform distributed optimizations on the query plan and avoid broadcast. Second, unlike other P2P systems, for XPeer to work, minimal human intervention is needed and no nodes with special administrative duties have to be permanently connected; nodes in the overlay network are themselves peers (thus can connect and disconnect at any time) entitled by the system to perform administrative tasks. Indeed, XPeer adopts self-administration policies, namely *root cloning protocols* and *network protocols*, to provide workload balance and robustness to network failures. The protocols will dynamically optimize the number of administrative nodes required by the overlay network to handle the query demand of the connected peers.

Publications: [CGMS07a] [SMGC04] [CGMS04b] [CGMS05] [CGMS07b]

2006 - ongoing *Data Models and Types for Digital Library Management Systems.*

Digital Library Management Systems (DLMS) are systems providing developers with the functionalities they require to develop Digital Library Systems (DLS), such as multimedia archives and institutional repositories. In the last few years, mainly due to the growth of the Digital Libraries and their integration and federation over the web, DLMS naturally evolved to support management of Compound Objects. Compound Objects conform to rather flexible data models, according to which DLS Information Spaces are formed by digital objects (e.g. documents, metadata) that can be combined by (typically named) relationships to form arbitrary information graphs. Existing DLMS platforms tend to support full flexibility (graph-stores) and renounce the benefits of static typing.

Paolo Manghi leads the DOROTY Lab (Marko Mikulicic, Michele Artini and Alessia Bardi) of the D-Lib group and is responsible of the research, design and development activities of the DOROTY project [Pr16]. DOROTY (*Digital Object Repository with Types*) is a DLMS whose novelty is to provide Digital Library Systems designers and developers with tools inspired by the traditional typed approach of Database Management Systems. Using DOROTY, DLS developers first declare the types of the objects to be manipulated and than realize applications counting on type-checking and optimized storage and access methodologies. DOROTY's data model offers a wide range of abstractions common to Digital Library

objects modeling. Examples are object collections, named relationships, object versioning, metadata descriptions, and others, as well as operations to manipulate them. Beyond recovering type correctness and data integrity properties, the adoption of types in DLMS foundations allows for the design of an efficient and optimized DLMS store, capable of exploiting type information to achieve best performances in object persistence and access.

As such DOROTY can ease the design and implementation of Digital Library Systems, such as multimedia archives and institutional repositories. The DLMS is currently used as back-end for the repository technologies developed in the DRIVER-II project [Pr7] (Enhanced Publication Service), the EFG project [Pr9] (European Film Information Space Service) and R2D2 [Pr8].

Publications: [CCM⁺09] [Man08c] [CCMP06] [CMP07] [CCM⁺08]

2006 - ongoing *Service-oriented infrastructures for the aggregation of distributed Digital Library Systems.*

In the last decade, there has been an increasing demand for integrating and sharing content and functionalities of independent Digital Library Systems. Typically, however, existing solutions target very specific issues and become not easily sustainable in time due to onerous realization and maintenance cost. Service-oriented infrastructure and resource sharing model (typical of the GRID approach) offer environments facilitating both functionality and content integration process, by establishing common participation and interaction practices and by distributing the cost among the stakeholders.

Since January 2006 Paolo Manghi moved to ISTI-CNR where he started research on Service Oriented Infrastructures for Digital Libraries. He leads the D-NET laboratory (Marko Mikulicic, Michele Artini, Alessia bardi and Claudio Atzori) at D-Lib research group, and he is responsible for the research, design, development, and deployment of the *D-NET software toolkit* [Pr17]. D-NET is the outcome of research activities inspired by “GRID resource discovery and sharing” principles and applied to service oriented architectures and Digital Library system definition and aggregation. In a running D-NET infrastructure data source and service providers can dynamically integrate (register) their resources, embedded as services, at run-time. Such services can in turn discover and interact with others registered to the infrastructure. The environment offers tools for handling autonomic behaviour of resources: special services automatically orchestrate and combine given resources typologies into different workflows, in order to achieve or guarantee an expected behaviour (e.g. replica management, availability of service). In particular, the service kits currently supported by D-NET permit the composition of applications whose services can collect and flexibly aggregate content from a federation of Digital Library systems (multimedia archives and institutional repositories) and supply end-users with portals to access and manipulate such content by means of advanced functionalities; e.g. search, collection management, user registration, query profiling, user communities, and others.

Due to its high flexibility and low uptake cost, D-NET is currently serving as software platform for a number of EU projects: DRIVER [Pr6] and DRIVER-II [Pr7] (which first funded its realization), the European Film Gateway [Pr9], OpenAIRE [Pr1], OpenAIREplus [Pr3], HOPE [Pr2], R2D2 [Pr8]. Furthermore, D-NET is the supporting platform for the national repository infrastructures of Spain-Recolecta (Consorcio Madrono, Spain) and Slovenia (National and University Library of Ljubljana), while it is under experimentation for Argentina, Bulgaria, Greece, and Poland.

Publications: [CCMP07b] [CCMP07a] [ABMM08] [ACC⁺08b] [ACC⁺08a] [dri07] [FHM⁺07] [efg07] [MPZ08a] [ACC⁺09] [IMP09b] [MMC⁺10b] [MMHP10] [ILM⁺10] [MPI09] [CKM11]

2009 - ongoing *Architectures for collaborative research environments*

Scholarly research involves the systematic study of information sources in order to establish facts and reach new conclusions. It encompasses survey, analysis, evaluation and creation as distinct phases that are performed iteratively or in parallel by accessing a range of local and remote resources. Throughout these activities scholars create collections of relevant work, ranging from publication references to new information acquired through experiments or correspondence with other scholars. We use the term reading list to refer to such collections. Existing software packages or web services for managing publication lists, like CiteULike, lack integration with researchers workflow, which may require access to both desktop and online resources. In this research we target the architecture and system design of ScholarLynk [Pr18], a desktop tagging tool that enables researchers to build and maintain reading lists across distributed data stores, in collaboration with other researchers.

Publications: [KMI⁺10a] [KMI⁺10b] [KMFH⁺10]

4.2 Funding Projects

Paolo Manghi was and is involved in the authoring, research, coordination and technical management activities of several funding projects, all mentioned in the previous section. In particular:

1. *Researcher* for the project BELIEF [Pr5].
2. *Author and conceiver* of the accepted project proposals: DRIVER-II Project [dri07][Pr7], EFG [efg07][Pr9], OpenAIRE [ope09][Pr1], HOPE [hop10][Pr2], R2D2 [r2d09][Pr8], OpenAIREplus [Pr3], EFG1914 [Pr4].
3. *Researcher and Technical Manager* for the projects DRIVER [Pr6], DRIVER-II [Pr7], EFG [Pr9], R2D2 [Pr8], HOPE [Pr2], OpenAIRE [Pr1], OpenAIREplus [Pr3], EFG1914 [Pr4].

In particular, he was later delegated as project consortium representative to meet the European Union reviewers/experts for the following EU proposal evaluation hearings:

25.06.2007 European Union - Call for Hearings (FP7, e-infrastructures): consortium scientific representative for project proposal DRIVER-II [Pr7].

16.05.2009 European Union - Call for Hearings (FP7, e-infrastructures): consortium scientific representative for the project proposal OpenAIRE [Pr1].

22-23.02.2011 European Union - Call for Hearings (FP7, e-infrastructures): consortium scientific representative for the project proposal OpenAIREplus [Pr3].

The following sections present the projects where he was directly involved, each accompanied by the project details: name, aims, research topics, scientific roles he covered, affiliated institution at the time of the project, and fundings granted. Projects are categorized in *ongoing* and *past*.

4.2.1 Ongoing projects

[Pr1] *OpenAIRE* (Open Access Infrastructure for Research in Europe): FP7 Combination of Collaborative projects and Coordination and Support Actions - Integrated Infrastructures Initiative project (I3) proposal (grant agreement: 246686, call: FP7-INFRASTRUCTURES-2009-1). The project is a key actor in the process of dissemination and uptake of the EU Open Access mandate. It aims at deploying and maintaining the OpenAIRE System, which enables the European Infrastructure for Open Access articles published under fundings from projects in the FP7. OpenAIRE offers a portal (www.openaire.eu) from which authors can deposit their articles, compliant Open Access repositories can register to be pro-actively harvested by the infrastructure, and users can search and access the articles or statistics on articles per project and per research area. [ope09] [CMB09] [CM10] [Man10] [MM10] [BEM⁺] [MM11]

Topics: service infrastructures for the aggregation of Digital Library systems and Digital Library Management Systems, automatic deposition in remote repositories, collaborative data curation, authority file management.

Roles: Proposal Author, Researcher, and Technical Manager;

Affiliated Institution: ISTI-CNR, role: project technical coordination (Dr. Donatella Castelli);

Funding: 660,000 EUR (project total: 4,170,000 EUR)

Duration: December 2009 - November 2012.

[Pr2] *HOPE* (Heritage of the Peoples Europe) (2010-2013) [hop10]: Best Practice Networks project (grant agreement: 250549 - FP7 EU eContentplus 2009). HOPE is a Best Practice Network of archives, libraries and museums of social and labor history institutions across Europe. It aims to improve access to the vast amount of highly significant but scattered digital collections on social history. It proposes to achieve this by promoting the adoption of standards and best practices for digital libraries amongst its partners, by ensuring that the metadata and the content become available through Europeana and by implementing a full scale discovery-to-delivery model. [SMMvdW10]

topics: service infrastructures for the aggregation of Digital Library systems;

Roles: Project WP Technical Coordinator, Proposal Author, Researcher, and Technical Manager;

Affiliated Institution: ISTI-CNR, role: project technical coordination (Dr. Paolo Manghi);

Funding: 330,000 EUR (project total: 2,700,000 EUR)

Duration: April 2010 - March 2014.

- [Pr3] *OpenAIREplus* (2nd-Generation Open Access Infrastructure for Research in Europe): FP7 Combination of Collaborative projects and Coordination and Support Actions - Integrated Infrastructures Initiative project (I3) proposal (grant agreement: 283595, call: FP7-INFRASTRUCTURES-2011-2). The project continues and extends the OpenAIRE System in the direction of openly share research data and link it to the relative publications.

Topics: service infrastructures for the aggregation of Digital Library systems and Digital Library Management Systems, automatic deposition in remote repositories, collaborative data curation, authority file management, research data.

Roles: Proposal Author, Researcher, and Technical Manager;

Affiliated Institution: ISTI-CNR, role: project technical coordination (Dr. Donatella Castelli);

Funding: 580,000 EUR (project total: 4,200,000 EUR)

Duration: December 2011 - June 2013.

- [Pr4] *EFG1914*: ICT Policy Support Programme - Pilot Type B (Grant agreement: 292276, Call: CIP-ICT-PSP-2011-5, 2.2. "Digitising content for Europeana). EFG1914 will digitise and make available 710 hours of film and 6.800 film-related items on the theme of World War I held by 21 European archives in 15 countries. The content will be made available through the EFG Portal and Europeana and, addition to that, in a special virtual exhibition dedicated to the content digitised in the project and to themes around WWI, film, the European film industries and their audiences in a decade of conflict and cataclysm. The project is supported by the Association des Cinmathques Europennes (ACE) and the Europeana Foundation itself.

Topics: service infrastructures for the aggregation of Digital Library systems, data curation, authority file management.

Roles: Researcher and Technical Manager;

Affiliated Institution: ISTI-CNR, role: project technical partner;

Funding: 580,000 EUR (project total: 4,200,000 EUR)

Duration: February 2012 - January 2014.

4.2.2 Past projects

In the following only the major projects he was involved with will be mentioned. However, he participated to several others, such as GRID.it (FIRB IT), DataX (IT), EPSRC (UK), Training Movement Grant (EC), NAPI project (Microsoft Research Cambridge).

- [Pr5] *BELIEF* (Bringing Europe's eElectronic Infrastructures to Expanding Frontiers): FP6 EC Specific Supporting Action project (grant agreement: 026500, Call: FP6-2004-IST-6 - 3.2.3 "Communication Network Development - eInfrastructure Consolidating Initiatives"). The aim of the project is to facilitate knowledge-exchange on eInfrastructures a one-stop home for public eInfrastructure documentation, e.g. project DoWs, project deliverables. This information will be readily accessible to BELIEF Community Members though the BELIEF Digital Library especially developed to provide a central repository for eInfrastructure Information.

Topics: Digital Library Systems and methodologies for document collection.

Roles: Researcher;

Affiliated Institution: ISTI-CNR, role: partner (Dr. Donatella Castelli);

Duration: November 2005 - December 2007

Fundings: 220,000 EUR (project total: 950,000 EUR)

- [Pr6] *DRIVER* (the Digital Repository Infrastructure Vision for European Research): FP6 EC Specific Targeted Research Project (grant agreement: IST-034047, Call: FP6-2005-IST-5 - 2.5.6.3 "Research Networking Testbeds"). The project's applicative goal is that of encouraging Open Access

business models among researchers and publishers by giving centralized access and visibility to this critical mass of publications already available from repositories in Europe. To this aim, the project funded the realization and deployment of the D-NET software toolkit and today operates the first European Open Access Repository Infrastructure. D-NET services are combined to form multiple distributed applications over the aggregation of content from a federation of institutional repositories (today 240, for about 2,500,000 digital object surrogates), i.e. archives containing any form of scientific output, including scientific/technical reports, working papers, pre-prints, articles and original research data. [MM06][Man06]

Topics: service infrastructures for the aggregation of Digital Library systems;

Roles: Researcher and Technical Manager;

Affiliated Institution: ISTI-CNR, role: project coordination (Dr. Donatella Castelli);

Funding: 430,000 EUR (project total: 2,700,000 EUR)

Duration: June 2006 - November 2007

- [Pr7] *DRIVER-II* (the Digital Repository Infrastructure Vision for European Research - Phase Two): FP7 EU Combination of Collaborative projects and Coordination and support actions project (grant agreement: 212147, call: INFRA-2007-1.2.1 - Scientific Digital Repositories). The project goal is to continue the operation of the DRIVER Infrastructure [Pr6] and to extend it with a Digital Library System for the management of “enhanced publications”, special compound objects grouping together articles and research data related with them. [Man08c] [MM08c] [MMSI08] [Man08d] [Man08b] [MS08] [Man08e] [MM08a] [MM08b] [Man08a]

Topics: service infrastructures for the aggregation of Digital Library Systems, Compound Object data models and DLMS, DOROTY [Pr16].

Roles: Proposal Author, Researcher and Technical Manager. *Goal:*

Affiliated Institution: ISTI-CNR, role: project coordination (Dr. Donatella Castelli);

Funding: 580,000 EUR (project total: 2,700,000 EUR)

Duration: December 2007 - November 2009

- [Pr8] *R2D2* [r2d09]: research project between Microsoft Corporation, Research Departments of Cambridge and Redmond (ref. Scarlet Schwiderski-Grosche), the D-Lib research group of ISTI-CNR (ref. Donatella Castelli) and the MaDgIK Laboratory (ref. Prof. Yannis Ioannidis) of the Dept. of Informatics and Telecommunications, National and Kapodistrian University of Athens. The projects aims at the realization of a Microsoft desktop tool called ScholarLynk, capable of enabling collaboration of researchers by exploiting support of service oriented infrastructures (D-NET software [Pr17]) and exploring new avenues in collaborative tagging systems.

Topics: tagging and collaborative systems for very large digital libraries, service infrastructures for the aggregation of Digital Library systems, Digital Library Management Systems.

Roles: Proposal Author, Researcher and Technical Manager;

Affiliated Institution: ISTI-CNR;

Funding: 100,000 EUR (project total 200,000 EUR).

Duration: October 2009 - September 2010.

- [Pr9] *EFG* (European Film Gateway) [efg07]: Best Practice Networks project (grant agreement: ECP 517006-EFG, call: FP7 EU eContentplus 2007). The project aims at deploying and maintaining a pan-European Film Gateway infrastructure, capable of gathering, aggregating and exposing film information and content available from Movie archives across European countries. The EFG online portal, will provide direct access to about 790,000 digital objects including films, photos, posters, drawings, sound material and text documents. Users will have the possibility to search (in multiple languages) and to browse through the digital objects. EFG aims at supporting technical and semantic interoperability between cinematographic archives and it will support the export to Europeana. EFG will also evaluate measures to be taken to deal with IPR issues. [EMS09] [MM11]

Research topics: service infrastructures and data models, multi-lingual search, authority file management;

Roles: Proposal Author, Researcher and Technical Manager;

Affiliated Institution: ISTI-CNR, role: project scientific coordination (Dr. Pasquale Savino);

Funding: 500,000 EUR (project total: 4,500,000 EUR);

Duration: September 2008 - August 2011.

4.3 Research projects

Paolo Manghi activities resulted in the following research projects, whose context and details can be found in section 4.1 *Topics*. Some of them were funded with targeted grants, others with a number of EC projects where the research products involved could be fruitfully used by real user communities. This is the case for D-NET [Pr17] and DOROTY [Pr16], developed, enhanced and used to support the real-case communities of DRIVER [Pr6], DRIVER-II [Pr7], EFG [Pr9], OpenAIRE [Pr1], and HOPE [Pr2] projects).

[Pr10] *Hippo* (High-level Internet Programming with Persistent Objects) (1997-1999). *Roles:* Researcher. *Affiliated Institution:* Department of Computer Science, Glasgow University (UK).

[Pr11] SNAQue (1998-2000). *Roles:* Researcher and Technical Manager. *Affiliated Institution:* Department of Computer Science, Strathclyde University (UK).

[Pr12] TeQuyLa (1999-2001). *Roles:* Researcher. *Affiliated Institution:* Dipartimento di Informatica, Università di Pisa.

[Pr13] TQL (2000-2001). *Roles:* Researcher. *Affiliated Institution:* Dipartimento di Informatica, Università di Pisa. [ACC⁺01]

[Pr14] microXQuery (2001-2003). *Roles:* Researcher. *Affiliated Institution:* Dipartimento di Informatica, Università di Pisa.

[Pr15] XPeer: P2P XML Database (2003-2006). *Roles:* Researcher and Technical Manager. *Affiliated Institution:* Dipartimento di Informatica, Università di Pisa. [CGMS06b]

[Pr16] DOROTY (Digital Object Repository with TYpes) (2009-ongoing). *Roles:* Researcher and Technical Manager. *Affiliated Institution:* Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa.

[Pr17] *D-NET Software Toolkit* (2007-ongoing). *Roles:* Researcher and Technical Manager. *Affiliated Institution:* Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa. [MM09]

[Pr18] *ScholarLynk* (2009-ongoing). *Roles:* Researcher and Technical Manager. *Affiliated Institution:* Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa, National and Kapodistrian University of Athens, Microsoft Research Cambridge. [MMLH10]

[Pr19] *PACE (Programmable Authority Control Engine)* (2010-ongoing). *Roles:* Researcher and Technical Manager. *Affiliated Institution:* Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Pisa. [MM11]

5 International cooperations

Paolo Manghi was and is involved in several international cooperations with scientists from academia and companies, through projects or because of common research interests. The most relevant are:

- From 01.10.1995 to 31.03.1996 he visited the PISA Research Group of Prof. R. Connor and Prof. R. Morrison at the Department of Computer Science of S.Andrews University (UK). The reason of the visit were studies and comparisons on programming methodologies for the orthogonally persistent languages *Napier88* (PISA research group, S.Andrews University) and *Fibonacci* (Fibonacci research group, University of Pisa).
- In 1997 he joined the research group led by Prof. R. Connor at Strathclyde University (UK) to work on the projects *Hippo* and *SNAQue*.

- In 2001/02 he joined the project *Network-Aware Programming and Interoperability* (NAPI), the result of a cooperation between Cambridge Microsoft Research (UK, references: Luca Cardelli and Bruno Quarta) and a selection of Italian Universities (Pisa, Bologna, Firenze, and Milano).
- In 2005 he cooperated with Redmond Microsoft Research to work on the project *Big Top*. Goal of the cooperation is the integration of *TQL* language within *HighWire*'s programming environment.
- In 2006 he started the DRIVER and DRIVER-II project activities, where he started a flourishing cooperation with the MaDgIK Laboratory of Prof. Yannis Ioannidis, Kapodistrian University of Athens, Dept. of Informatics and Telecommunications, Wolfram Horstmann from the University of Bielefeld and Wojtek Sylwestrzak from ICM research center of Warsaw (Poland).
- In the period from 05.05.2008 to 11.05.2008 2008 he visited Prof. Yannis Ioannidis at Kapodistrian University of Athens to start a work on data models and types for digital libraries.
- In 2008 he became a working group member of *DL.org Expertise* an activity of the EU project DL.org www.dlorg.eu. The working group aims at defining best practices, models and standards on interoperability for Digital Libraries with respect Content. *Publications*: [TCC⁺10]
- In 2009 he joined the Core Expert Group of the major Europeana EU Project (<http://www.europeana.eu>), which aims at building a common multi-lingual access point to Europe's distributed cultural and scientific heritage, including digital content from all types of heritage institutions (archives, libraries, museums and audio-visual collections). The group works at refining and identifying the specifications of the data model and architecture of Europeana.
- In 2009 he started a cooperation with Microsoft Research, departments of Cambridge (ref. Fabrizio Gagliardi, Scarlet Schwiderski-Grosche, Natasa Milic-Frayling and Gabriella Kazai) and Redmond (ref. Alex Wide and Lee Dirks), on integration between Microsoft Desktop and the D-NET-based DRIVER infrastructure [Pr6][Pr7]. The outcome of the cooperation is the project R2D2 [Pr8] and the software tool ScholarLynk [MMLH10].

6 Conferences organization and invited talks

Paolo Manghi is the organizer and editor of the Very Large Digital Libraries international workshop, at its fourth edition this year and so far held in conjunction with the Theory and Practice of Digital Libraries (formerly European Conference on Digital libraries). The workshop is the outcome of the several activities and international cooperations developed in the area of infrastructures and content management for Digital Libraries. For similar reasons, he was invited to present his research in the context of workshops, conferences and work-groups.

Conference Organization and Program Committees

- Production Editor** 8th International Workshop on Database Programming Languages (DBPL 2001), Sept. 2001, Rome, Italy. Chair: Prof. Giorgio Ghelli.
- Organizer and Editor** First International Workshop on Very Large Digital Libraries (ECDL 2008), Sept. 2008, Aarhus, Denmark (SIGMOD workshop report [MPZ08b] [MPZ08a])
- Organizer and Editor** Second International Workshop on Very Large Digital Libraries (ECDL 2009), Sept. 2009, Corfu, Greece [IMP09a] [IMP09b] [MPI09]
- Organizer and Editor** Third International Workshop on Very Large Digital Libraries (ECDL 2010), Sept. 2010, Glasgow, UK. [IMP10]
- Organizer and Editor** Fourth International Workshop on Very Large Digital Libraries (TPDL 2011), Sept. 2011, Berlin, Germany. [CIM11b] [CIM11a]
- Program Committee** . BooksOnline2010 Workshop, in conjunction with CIKM Conference 2011. Chair: Gabriella Kazai (Microsoft Research, UK).
- Program Committee** BooksOnline2011 Workshop, in conjunction with CIKM Conference 2011. Chair: Gabriella Kazai (Microsoft Research, UK).
- Program Committee** Metadata and Semantics Research Conference 2011, OAR Track. Chair: Elena Garca-Barriocanal (University of Alcal); co-chairs: Zeynel Cebeci (ukurova University, Turkey), Mehmet C. Okur, (Yasar University, Turkey).

Program Committee World Wide Web Conference 2012, Demo Track. Chair: Antoine Isaac, Daniel Muller, and Bebo White. Lyon, France.

Program Committee Scientific and Statistical Database Management Conference 2012. Chair: Mike Chatzopoulos. Chania, Crete, Greece.

Reviewer activities

Database and Programming Language Conference, Springer Verlag Reviewer in 2001. Organizer: Giorgio Ghelli.

Data and Knowledge Engineering Journal, Elsevier Reviewer in 2004. Editor in Chief: Peter Chen.

International Journal on Digital Libraries, Springer Reviewer in November 2008. Associate Editor: Carol Ann Peters.

Open Repositories Conference Reviewer in 2010. Chair: Dr. Wolfram Horstmann.

Other reviewing activities were carried out for conferences and workshops such as Italian Symposium on Database Systems (SEBD) and Italian Research Conference on Digital Libraries (IRCDL).

Invited talks

20-22.08.1996: invited speaker at EC-US Workshop on Persistence (Kinloch Rannoch, Scotland). Presentation title: *Programming methodologies for Persistent Programming Languages*.

17-19.05.1998: invited speaker at Pastel Workshop (Fort Williams, Scotland). Presentation title: *On the Unification of Persistent Programming and the World Wide Web*.

07-09.12.1998 invited speaker at Workshop on Persistence (Kinloch Rannoch, Scotland). Presentation title: *Interning Never Externed Data*.

07-09.12.2001 invited speaker at First Workshop on Network-Aware Programming and Interoperability (NAPI) at Microsoft Research (Cambridge, England).

11-12.02.2002 invited speaker at Second Workshop on Network-Aware Programming and Interoperability (NAPI) in San Miniato, Pisa.

27.05.2005 invited speaker at *BigTop* project Workshop (MS Research), Dipartimento di Informatica, University of Bologna. Presentation title: *XPeer: a P2P XML database system*

29.05.2007 invited speaker at the DELOS Network of Excellence School on Digital Libraries, Settignano, Florence, Italy. Presentation title: *The DRIVER Repository Infrastructure*.

25.06.2007 invited speaker at SURF Foundation - Utrecht, The Netherlands. Presentation title: *Compound Objects*.

18.02.2008 invited speaker at the DRIVER Open Summit, Goettingen, Germany. Presentation title: *The DRIVER Infrastructure architecture*.

02.10.2008 invited speaker at TrebleCLEF Workshop, Wolfsburg, Swiss. Presentation title: *The DRIVER Infrastructure architecture: multilinguality issues*.

16.11.2008 invited speaker at SPARC Digital Repositories Meeting, Baltimore, USA. Presentation title: *Building Sustainable Aggregative Digital Library System*.

13-14.01.2009 invited speaker at EuropeanaLocal Knowledge Sharing Workshop, Den Haag, The Netherlands. Presentation title: *The DRIVER Infrastructure*.

16-18.03.2009 invited speaker at DRIVER/JISC Digital Repositories Workshop, Amsterdam, The Netherlands. Presentation title: *Typed Compound Object Models for Digital Libraries*.

03.09.2009 invited speaker at Società Internazionale per lo Studio del Medioevo Latino (SISMEL), Certosa del Galluzzo, Florence. Presentation title: *Aggregating content from heterogeneous metadata data sources*.

18.02.2010 invited speaker at European Science Foundation meeting, ILC-CNR, Pisa. Presentation title: *D-NET Software Toolkit and the DRIVER Infrastructure experience*. Ref: Dr. Andrea Bozzi, director of the Istituto di Linguistica Computazionale, CNR.

- 06.07.2010** invited speaker at Birth of Feather session *Handling of compound objects from an e-Infrastructure point of view* at the Open Repository conference, Madrid, Spain, 6-9 of July, 2010. Ref: Jochen Schirrwagen, University of Bielfeld, Germany.
- 07.07.2010** invited speaker at Birth of Feather session *Interoperability in Digital Libraries* at the Open Repository conference, Madrid, Spain, 6-9 of July, 2010. Ref: Giuseppina Vullo, University of Glasgow, Scotland.
- 04.05.2011** invited speaker at the INFISO-RTD Open Access coordination workshop, EC Offices, Brussels, Belgium. Presentation title: *Data acquisition in OpenAIRE*. Ref: Carlos Morais Pires (Head of Scientific Data Infrastructure sector, e-Infrastructures Unit, 7th Framework Programme).
- 22-23.05.2011** invited speaker at the Integrating Research Information CRIS+OAR workshop (EuroCRIS and CNR), CNR offices, Rome, Italy. Presentation title: *Architectures based on Repositories: The OpenAIRE experience*. Ref: Keith Jeffery, Daniela Luzi, Anne Asserson.
- 11-33.09.2011** invited speaker at the CERIF-CRIS and Repositories workshop (EuroCRIS), Brussels, Belgium. Presentation title: *Introducing CRIS Information in Repository Federations – The OpenAIRE Project Experience*. Ref: Keith Jeffery, Anne Asserson.
- 12-13.10.2011** invited speaker at the e-IRG workshop, session “Data infrastructure and data Management”, Poznan, Poland. Presentation title: *OpenAIREplus project: Towards an European Open Access infrastructure for Research*. Ref: Norbert Meyer.

7 Teaching and research supervision

Since he moved to the University of Pisa, Paolo Manghi started a regular teaching activity, which only recently he had to reduce due to growing research and project coordination responsibilities. His major experience is in teaching database design principles, but also other topics were touched, such as architectures and programming languages and methodologies. Teaching has naturally evolved in the years into MSC students and post-doc students supervision.

Member of CNR commissions for public selection of personnel

- “**Commissione esaminatrice di pubblica selezione**” **prot. 953, 23.04.2009** Selection for 1 position as temporary research fellow (“assegnista di ricerca”) to work on the project DRIVER-II. Winner of selection: Alessia Bardi.
- “**Commissione esaminatrice di pubblica selezione**” **prot. 2210, 28.09.2010** Selection for 1 position as temporary research fellow (“assegnista di ricerca”) to work on the project D4Science-II. Winner of selection: Sandro La Bruzzo.

Post-doc supervision

D-Lib workgroup on “content for digital libraries” The workgroup numbers the following research staff: Marko Mikulicic, Alessia Bardi, Cristina Tang, Claudio Atzori, Sandro La Bruzzo e Michele Artini. Its main focus is on studies, research and development in the area of Digital Library Management Systems, namely provenance information and data models for compound objects.

Student supervision

- Tirocinio 2007** Emanuele Cavarretta, Dipartimento di Informatica, Università di Pisa: *Regrouper (research group publisher)*
- Laurea specialistica 2008-2009 - MSc thesis** Alessia Bardi, Dipartimento di Informatica, Università di Pisa: *DOROTY: Digital Object RepOsitory with Types* (for more information refer to section research activities of this CV);
- Laurea specialistica 2009-2010 - MSc thesis** Sandro La Bruzzo, Dipartimento di Informatica, Università di Pisa: *Implementazione di interfacce OAI per l’esportazione di Compound Objects nel sistema DOROTY*; work on design and implementation of standard data-export interfaces for the typed data model of the DOROTY DLMS (for more information see research activities section of this CV).

University, Master, and Post-Doc lecturer

- December 2000** Course: *Basi di Dati*, Master Web & Wireless. Organizers: University of Pisa and Vodafone (ex Omnitel). Duration: 1 day.
- December 2001** Course: *Applicazioni Web per Basi di Dati*, Web & Net Master. Organizers: Pisa Province Council, University of Pisa (reference: Dr. Laura Ricci), and Qualital Consortium. Duration: 2 weeks.
- January 2002** Course: *Basi di Dati*, IFTS course *Tecnico Informatico per il supporto al commercio elettronico e ai servizi informativi territoriali*. Organizers: Regione Toscana and University of Pisa (reference: Prof. Dino Pedreschi). Duration: semester.
- May 2004** Course: *Database, MASTER SIT*. Polo Scientifico e Tecnologico dell'Area Livornese Srl, Livorno. Duration: 2 weeks.
- December 2004** Course: *IT4PS - Access*. Organizers: Dipartimento di patologia sperimentale biotecnologie mediche, infettivologia e epidemiologia, University of Pisa, CRUI (Conferenza dei Rettori delle Università Italiane), and AICA (Associazione Italiana per l'Informatica e il Calcolo Automatico). Duration: 2 weeks.
- March 2005** Course: Master "Open Source". Seminars: *Le Interrogazioni SQL e Ambienti per la definizione delle interrogazioni*. Dipartimento di Informatica, University of Pisa. Duration: 2 days (11.03.2005, 16.03.2005).
- September 2006** Course: Master "Open Source". Seminars: *Il ruolo della progettazione nello sviluppo delle Basi di Dati e Il Linguaggio SQL*. Dipartimento di Informatica, University of Pisa. Duration: 2 days (08.09.2006, 15.09.2006).
- September 2007** Course: Master "Open Source". Seminars on *SQL Data Definition Language*. Dipartimento di Informatica, University of Pisa. Duration: 3 days (14.09, 19.10, 11.11.2007).
- October 2007** Course: *On Teaching Databases*, SSIS in Informatica (Teachers School of Specialization), Dipartimento di Informatica di Pisa. Duration: 1 month.
- August 2008** Module *Put Yourself in the DRIVER's Seat - Practical Training for Building a European Repository Network* at the Summer School *Digital Libraries la Carte 2008*, Tilburg University, The Netherlands. Duration: 1 day (28.08.2008).
- April 2009** Course: *On Teaching Databases*, SSIS in Informatica (Teachers School of Specialization), Dipartimento di Informatica di Pisa. Duration: 2 months (09.04-06.06.2009).
- June 2009** Module *D-NET Technology* of the DRIVER Tutorial at the CERN workshop on Innovations in Scholarly Communication, OAI6 Conference, University of Geneva, Geneva, 16-19 of June, 2009. Duration: 1 day (17.06.2009).
- October 2010** Lecture: *Content Interoperability in Digital Libraries* at the *DL.org Autumn School*, National and Kapodistrian University of Athens, Athens, Greece, 3-8 of October, 2010. Duration: 2 hours.

University teaching assistant (contratti di supporto alla didattica)

- March 2000** Course: *Basi di Dati - Laboratorio Oracle*, Laurea in Informatica, Dipartimento di Informatica, University of Pisa. Duration: semester (from 20.03.2000).
- October 2000** Course: *Architetture degli Elaboratori I - Operative Systems*, Diploma in Informatica, Dipartimento di Informatica, University of Pisa. Duration: semester (from 23.10.2000).
- March 2001** Course: *Basi di Dati - Laboratorio Oracle*, Laurea in Informatica, Dipartimento di Informatica, University of Pisa. Duration: semester (from 12.03.2001).
- March 2002** Course: *Basi di Dati - Laboratorio Oracle*, Laurea in Informatica, Dipartimento di Informatica, University of Pisa. Duration: semester (from 18.03.2002).
- October 2002** Course: *Laboratorio di Introduzione alla Programmazione - Java*, Laurea in Informatica, Dipartimento di Informatica, University of Pisa. Duration: semester (from 16.10.2002).
- March 2003** Course: *corso Basi di Dati - Laboratorio Oracle*, Laurea in Informatica, Dipartimento di Informatica, University of Pisa. Duration: semester (from 01.03.2003).
- October 2003** Course: *Laboratorio di Introduzione alla Programmazione - Java*, Laurea in Informatica, Dipartimento di Informatica, University of Pisa. Duration: semester (from 20.10.2003).

March 2004 Course: *Basi di Dati - Laboratorio Oracle*, Laurea in Informatica, Dipartimento di Informatica, University of Pisa. Duration: semester (from 01.03.2004).

September 2005 Course: *Laboratorio di Introduzione alla Programmazione - Java*, Laurea in Informatica, Dipartimento di Informatica, University of Pisa. Duration: semester (from 29.09.2005).

April 2006 Course: *Informatica II*, Corso di Laurea in Comunicazione Pubblica Sociale e Impresa, Facoltà di Lettere, University of Pisa. Duration: semester (from 05.04.2006).

Professional teaching

November 2001 Module: *Basi di Dati - Analisi e Progettazione*. Specialization course *WF1 - Amministratore di Rete e Sistemista*. Organizer: Lucense S.p.A., Lucca. Duration: 1 week (01.11-05.09.2001).

December 2001 Module: *Basi di Dati - Analisi e Progettazione*. Specialization course *WF2 - Multimedia developer per Beni Culturali*. Organizer: Lucense S.p.A., Lucca. Duration: 1 week (03.12-05.09.2001).

July 2002 Module: *Programmi informatici Office*. *Olimpo* project. Organizers: Italian CCIAA and AsseforCamere S.r.l., Firenze. Duration 1 day (03.07.2002).

September 2002 Module: *Excel*. Project *POR Formazione innovativa del Management per la cultura*. Organizers: Province of Florence and Asseforcamere S.r.l., Firenze. Duration: 1 day (13.09.2002).

October 2002 Module: *MS Access*. Project *POR Formazione innovativa del Management per la cultura*. Organizers: Province of Florence and Asseforcamere S.r.l., Firenze. Duration: 1 day (29.10.2002).

November 2002 Module: *Principi di Database*. Specialization course *SIT PUBLIC*. Organizer: Società per il Polo Scientifico e Tecnologico dell'Area Livornese Srl, Livorno. Duration: 1 day (11.11.2002).

January 2003 Module: *HTML*. Specialization course: *UMTS*. Organizer: Képos S.r.l., Modena. Duration: 1 week (08.01-17.01.2003).

February 2003 Module: *Applicazioni Web - ASP*. Specialization course *UMTS*. Organizer: Képos S.r.l., Modena. Duration: 1 week (03.02-07.02.2003).

February 2003 Module: *Microsoft Access e Power Point*. Specialization course *Addetti all'orientamento attività imprenditoriali e Operatori sportelli territoriali*. Organizers: Unione Regionale Toscana delle Camere di Commercio and AsseforCamere S.r.l., Firenze. Duration: 1 day (10.02.2003).

May 2004 Module: *Database*. Specialization course *MASTER SIT*. Organizer: Società per il Polo Scientifico e Tecnologico dell'Area Livornese Srl, Livorno. Duration: 1 week (03.05-07-05.2004).

8 Publications

Thesis

- [Man96] Paolo Manghi. *Aspetti Linguistici della Costruzione di Applicazioni Persistenti*. Master's thesis, Dipartimento di Informatica Università di Pisa, July 1996
- [Man01] Paolo Manghi. *Extracting Typed Values from Semistructured Databases*. PhD thesis, Dipartimento di Informatica, University of Pisa, 2001. Supervisors: Giorgio Ghelli and Richard Connor

Books and book chapters

- [AAV⁺05] Brogi A., Martinelli A., Gervasi V., Manghi P., and Pacini G. Fabrizio A. *Il Foglio Elettronico per Medicina e Farmacia*. McGraw Hill, May 2005
- [PAA⁺05] Manghi P., Brogi A., Martinelli A., Gervasi V., Fabrizio A., and Pacini G. *Basi di Dati per Medicina e Farmacia*. McGraw Hill, November 2005

- [CKM11] Leonardo Candela, Akrivi Katifori, and Paolo Manghi. e-infrastructures. In Christian Meier zu Verl and Wolfram Horstmann, editors, *Studies on Subject-Specific Requirements for Open Access Infrastructure*, pages 125–164. Universitätsbibliothek Bielefeld, 2011

Journals

- [ADG⁺01] Albano A., Colazzo D., Ghelli G., Manghi P., and Sartiani C. A Typed Text Retrieval Query Language for XML Documents. *JASIST Special Topic Issue on XML and Information Retrieval*, 2001
- [SML⁺02] Fabio Simeoni, Paolo Manghi, David Lievens, Richard C.H. Connor, and Steve Neely. An Approach to High-Level Language Bindings for XML. *Elsevier Journal on Information and Software Technology, Special Issue on Objects, <XML> and Databases*, 44(4):217–228, 2002
- [SLCM03] Fabio Simeoni, David Lievens, Richard C.H. Connor, and Paolo Manghi. Language bindings to XML. *IEEE Internet Computing Journal*, 7(1):19–27, 2003
- [CGMS06a] Dario Colazzo, Giorgio Ghelli, Paolo Manghi, and Carlo Sartiani. Static Analysis for Path Correctness of XML queries. *Journal of Functional Programming (JFP)*, 16(4-5):621–661, 2006
- [FHM⁺07] Martin Feijen, Wolfram Horstmann, Paolo Manghi, Mary Robinson, and Rosemary Russell. DRIVER: Building the Network for Accessing Digital Repositories across Europe. *Ariadne*, 53, 2007
- [ACC⁺09] Michele Artini, Leonardo Candela, Donatella Castelli, Paolo Manghi, Marko Mikulicic, and Pasquale Pagano. Aggregative Digital Library Systems in the DRIVER Infrastructure. *World Digital Libraries Journal*, 2(2):113–130, December 2009
- [MMC⁺10b] Paolo Manghi, Marko Mikulicic, Leonardo Candela, Donatella Castelli, and Pasquale Pagano. Realizing and Maintaining Aggregative Digital Library Systems: D-NET Software Toolkit and OAIster System. *D-Lib Magazine*, 16(3/4), March/April 2010
- [MMHP10] Paolo Manghi, Natalia Manola, Wolfram Horstmann, and Dale Peters. An Infrastructure for Managing EC Funded Research Output – The OpenAIRE Project. *International Journal on Grey Literature (TGJ)*, 6(1), Spring 2010

Conferences

- [CGM96] Richard C.H. Connor, Giorgio Ghelli, and Paolo Manghi. Modules and Type Abstraction in Persistent Systems. In *7th International Workshop in Persistent Object Systems, Persistent Object Systems (POS) - Principles and Practice*, Cape May, New Jersey, May 1996
- [CGM96] Richard C.H. Connor, Giorgio Ghelli, and Paolo Manghi. Modules and Type Abstraction in Persistent Systems. In *7th International Workshop in Persistent Object Systems, Persistent Object Systems (POS) - Principles and Practice*, Cape May, New Jersey, May 1996
- [CMS98] Richard C.H. Connor, Paolo Manghi, and Keith Sibson. On the Unification of Persistent Programming and the World Wide Web. In *Workshop on the Web and Data Bases (WebDB98) held in conjunction with EDBT'98*, Valencia, Spain, 1998. Springer Verlag, LCNS series Vol. 1590
- [ACG⁺00] Antonio Albano, Dario Colazzo, Giorgio Ghelli, Paolo Manghi, and Carlo Sartiani. A Type System for Querying XML Documents. In *SIGIR00-XML ACM Workshop held in conjunction with 23rd International SIGIR Conference on Research and Development in Information Retrieval*, Athens, Greece, 2000
- [CGA⁺02a] Giovanni Conforti, Giorgio Ghelli, Antonio Albano, Dario Colazzo, Paolo Manghi, and Carlo Sartiani. The Query Language TQL. In *5th International Workshop on Web and Data Bases (WebDB02) in conjunction with ACM SIGMOD 2002*, pages 13–18, 2002
- [MSLC02] Paolo Manghi, Fabio Simeoni, David Lievens, and Richard C. H. Connor. Hybrid Applications over XML: Integrating the Procedural and Declarative Approaches. In *Fourth ACM CIKM International Workshop on Web Information and Data Management (WIDM)*, pages 9–14, SAIC Headquarters, LcLean, Virginia, USA, 2002. ACM

- [CGMS04a] Dario Colazzo, Giorgio Ghelli, Paolo Manghi, and Carlo Sartiani. Types for Path Correctness of XML Queries. In *Proceedings of the Ninth ACM SIGPLAN International Conference on Functional Programming (ICFP)*, pages 126–137, Snow Bird, UT, USA, September 2004. ACM
- [CCMP07b] Leonardo Candela, Donatella Castelli, Paolo Manghi, and Pasquale Pagano. OAI-Aggregator Services for Enhanced Cross-Repository Services. In *Post-proceedings of the Third Italian Research Conference on Digital Library Systems (IRC DL)*, pages 61–66, Padua, Italy, January 2007. DELOS: a Network of Excellence on Digital Libraries
- [CCMP07a] Leonardo Candela, Donatella Castelli, Paolo Manghi, and Pasquale Pagano. Enabling Services in Knowledge Infrastructures: The DRIVER Experience. In *Post-proceedings of the Third Italian Research Conference on Digital Library Systems (IRC DL)*, pages 71–77, Padua, Italy, January 2007. DELOS: a Network of Excellence on Digital Libraries
- [CGMS07a] Giovanni Conforti, Giorgio Ghelli, Paolo Manghi, and Carlo Sartiani. Scalable Query Dissemination in XPeer. In *Eleventh International Database Engineering and Applications Symposium (IDEAS 2007)*, pages 199–207, Banff, Alberta, Canada, September 2007. IEEE Computer Society
- [ABMM08] Michele Artini, Federico Biagini, Paolo Manghi, and Marko Mikulicic. OAI-Publishers in Repository Infrastructures. In *Post-proceedings of the Forth Italian Research Conference on Digital Library Systems (IRC DL)*, pages 93–98, Padua, Italy, January 2008. DELOS: an Association for Digital Libraries
- [ACC⁺08a] Michele Artini, Leonardo Candela, Donatella Castelli, Paolo Manghi, Marko Mikulicic, and Pasquale Pagano. Sustainable Digital Library Systems over the DRIVER Repository Infrastructure. In *Research and Advanced Technology for Digital Libraries, 12th European Conference (ECDL)*, volume 5173 of *Lecture Notes in Computer Science*, pages 227–231, Aarhus, Denmark, 2008. Springer
- [ACC⁺08b] Massimiliano Assante, Leonardo Candela, Donatella Castelli, Luca Frosini, Lucio Lelii, Paolo Manghi, Andrea Manzi, Pasquale Pagano, and Manuele Simi. An Extensible Virtual Digital Libraries Generator. In *Research and Advanced Technology for Digital Libraries, 12th European Conference (ECDL)*, volume 5173 of *Lecture Notes in Computer Science*, pages 122–134, Aarhus, Denmark, 2008. Springer
- [CCM⁺09] Leonardo Candela, Donatella Castelli, Paolo Manghi, Marko Mikulicic, and Pasquale Pagano. On Foundations of Typed Data Models for Digital Libraries. In *Fifth Italian Research Conference on Digital Libraries (IRC DL)*, Padua, Italy, 2009
- [TCC⁺10] Cristina Tang, Donatella Castelli, Leonardo Candela, Paolo Manghi, Pasquale Pagano, and Costantino Thanos. An Event-centric Provenance Model for Digital Libraries. In *Sixth Italian Research Conference on Digital Library Management Systems (IRC DL 2010)*, 2010
- [ILM⁺10] Katerina Iatropoulou, Antonis Lebesis, Paolo Manghi, Natali Manola, and Marko Mikulicic. On Constructing Repository Infrastructures - The D-NET Software Toolkit. In *Proceedings of the Fifth International Conference on Open Repositories*, Madrid, Spain, July 2010
- [MMC⁺10a] Paolo Manghi, Marko Mikulicic, Leonardo Candela, Michele Artini, and Alessia Bardi. General-Purpose Digital Library Content Laboratory Systems. In *Proceedings of the 14th European Conference on Digital Libraries*, Glasgow, UK, September 2010
- [KMI⁺10a] Gabriella Kazai, Paolo Manghi, Katerina Iatropoulou, Tim Haughton, Marko Mikulicic, Antonis Lempesis, Natasa Milic-Frayling, and Natalia Manola. Architecture for a Collaborative Research Environment based on the notion of Reading List Sharing. In *Proceedings of the 14th European Conference on Digital Libraries*, Glasgow, UK, September 2010
- [MM11] Paolo Manghi and Marko Mikulicic. PACE: A General-purpose Tool for Authority Control. In *Proceedings of the 5th Metadata and Semantics Research Conference*, Izmir, Turkey, October 2011

Relevant project reports

- [MM06] Paolo Manghi and Natalia Manola. Functional specification document. Deliverable, DRIVER Project, proposal number: IST-034047, August 2006
- [Man06] Paolo Manghi. Architectural specification document. Deliverable, DRIVER Project, proposal number: IST-034047, August 2006
- [Man08c] Paolo Manghi. Compound object model specification document. Deliverable, DRIVER-II Project, proposal number: 212147, September 2008
- [MM08c] Paolo Manghi and Marko Mikulicic. Service layer expansion. Deliverable, D5.2, 2008. DRIVER II. Deliverable D5.2, 2008
- [MMSI08] Paolo Manghi, Marko Mikulicic, Jochen Schirrwagen, and Marek Imialek. Aggregation advancement specification. Deliverable, D8.2, 2008. DRIVER II. Deliverable D8.2, 2008
- [Man08d] Paolo Manghi. D1.3 detailed activity plan report. Deliverable, 1.3, 2008. DRIVER II. Deliverable 1.3, 2008
- [Man08b] Paolo Manghi. Compound object model integration report d8.4. Deliverable, D8.4, 2008. DRIVER II. Deliverable D8.4, 2008
- [MS08] Paolo Manghi and Wojtek Sylwestrzak. Enabling services enhancement specification. Deliverable, D7.1, 2008. Driver II. Deliverable D7.1, 2008
- [Man08e] Paolo Manghi. Information space report. Deliverable, D5.1, 2008. DRIVER II. Deliverable D5.1, 2008
- [MM08a] Paolo Manghi and Marko Mikulicic. Monitoring tools specification. Deliverable, D7.3, 2008. DRIVER II. Deliverable D7.3, 2008
- [MM08b] Paolo Manghi and Marko Mikulicic. Overall research design report. Deliverable, D7.2, 2008. DRIVER II. Deliverable D7.2, 2008
- [Man08a] Paolo Manghi. Software release plan. Deliverable, D6.1, 2008. DRIVER II. Deliverable D6.1, 2008
- [EMS09] Georg Eckes, Paolo Manghi, and Pasquale Savino. Functional analysis of driver customisation. Deliverable, D4.1, 2009. EFG. Deliverable D4.1, 2009
- [MM10] Paolo Manghi and Marko Mikulicic. Openaire specification and release plan. Deliverable, D6.1, 2010. OpenAIRE. Deliverable D6.1, 2010
- [Man10] Paolo Manghi. Openaire data model specification. Deliverable, D5.1, 2010. OpenAIRE. Deliverable D5.1, 2010
- [SMMvdW10] Sjoerd Siebinga, Paolo Manghi, Mario Mieldijk, and Titia van der Werf. High-level Design of the HOPE Architecture. Deliverable, D2.1, 2010. HOPE. Deliverable D2.1, 31st August 2010
- [BEM⁺] Magchiel Bijsterbosch, Mikael Karstensen Elbek, Paolo Manghi, Natalia Manola, Jochen Schirrwagen, and Maurice Vanderfeesten. Technical report

Project proposals

- [dri07] DRIVER-II Project Proposal. Project proposal, Accepted for 7FP Call on Research Infrastructures INFRA-2007-1.2.1 - Scientific Digital Repositories, Combination of Collaborative projects and Coordination and Support Actions. Two-year project, started in December 2007 and ended in November 2009, grant agreement: 212147, 2007
- [efg07] European Film Gateway Project Proposal. Project proposal, Accepted by Best Practice Network funded by the European Commission under the eContent*plus* programme, as part of the i2010 policy. Three year project started in September 2008, grant agreement: ECP 517006 - EFG, 2007

- [ope09] Open Access Infrastructure for Research in Europe Proposal. Project proposal, Accepted by the European Commission under the FP7-INFRASTRUCTURES-2009-1. Combination of Collaborative projects and Coordination and Support Actions. Three-year project, started in December 2009, 2009
- [ope11] 2nd-Generation Open Access Infrastructure for Research in Europe. Project proposal, Accepted by EC FP7 Combination of Collaborative projects and Coordination and Support Actions. Two-year project, start date December 2011, 2011
- [efg12] EFG1914. Project proposal, Accepted by EC ICT Policy Support Programme - Pilot Type B. Two-year project, start date February 2012, 2012

Workshops

- [CLM+01] Richard C.H. Connor, David Lievens, Paolo Manghi, Steve Neely, and Fabio Simeoni. Extracting Typed Values from XML Databases. In *OOPSLA Workshop on Objects, <XML> and Databases*, Tampa Bay, Florida, USA, 2001
- [SMGC04] Carlo Sartiani, Paolo Manghi, Giorgio Ghelli, and Giovanni Conforti. XPeer: A Self-Organizing XML P2P Database System. In *EDBT Workshops*, volume 3268 of *Lecture Notes in Computer Science*, pages 456–465, Heraklion, Crete, Greece, 2004. Springer
- [CGMS04b] Giovanni Conforti, Giorgio Ghelli, Paolo Manghi, and Carlo Sartiani. A Self-organizing XML P2P Database System. In *Proceedings of the Twelfth Italian Symposium on Advanced Database Systems (SEBD)*, pages 394–401, S. Margherita di Pula, Cagliari, Italy, June 2004
- [CGMS05] Dario Colazzo, Giorgio Ghelli, Paolo Manghi, and Carlo Sartiani. Types for Path Correctness of XML Queries. In *Proceedings of the Thirteenth Italian Symposium on Advanced Database Systems (SEBD)*, pages 264–271, Brixen-Bressanone, Italy, 2005
- [CCMP06] Leonardo Candela, Donatella Castelli, Paolo Manghi, and Pasquale Pagano. Typing OpenDLib Repository Service: Strengths of an Information Object Type Language. In *1st European Workshop on the use of Digital Object Repository Systems in Digital Libraries (DORS DL), in conjunction with ECDL2006*, Alicante, Spain, 2006
- [CGMS07b] Giovanni Conforti, Giorgio Ghelli, Paolo Manghi, and Carlo Sartiani. Scalable Query Dissemination in XPeer. In *Proceedings of the 1st Workshop on Emerging Research Opportunities for Web Data Management (EROW 2007) Collocated with the 11th International Conference on Database Theory (ICDT 2007)*, volume 229 of *CEUR Workshop Proceedings*. CEUR-WS.org, January 2007
- [CMP07] Leonardo Candela, Paolo Manghi, and Pasquale Pagano. An Architecture for Type-based Repository Systems. In *Proceedings of the Second Workshop on Foundations of Digital Libraries, in conjunction with ECDL2007*, Budapest, Hungary, September 2007. DELOS
- [MCP10] Paolo Manghi, Leonardo Candela, and Pasquale Pagano. Interoperability Patterns in Digital Library Systems Federations. In *Proceedings of the Second DL.org Workshop on Making Digital Libraries Interoperable: Challenges and Approaches, in conjunction with ECDL 2010*, Glasgow, Scotland (UK), September 2010. ISTI-CNR

Proceedings, demos and posters

- [CGA+02b] Giovanni Conforti, Giorgio Ghelli, Antonio Albano, Dario Colazzo, Paolo Manghi, and Carlo Sartiani. The Query Language TQL - Demo Presentation. In *X Convegno nazionale su Sistemi Evoluti per Basi di Dati (SEBD)*, pages 427–431, Portoferraio, Italy, June 2002
- [MPZ08b] Paolo Manghi, Pasquale Pagano, and Pavel Zezula, editors. *Proceedings of the First Workshop on Very Large Digital Libraries (VLDL 2008), in conjunction with ECDL2008*, Aarhus, Denmark, September 2008. Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Italy, DELOS Association for Digital Libraries
- [MPZ08a] Paolo Manghi, Pasquale Pagano, and Pavel Zezula. First workshop on Very Large Digital Libraries – VLDL 2008. *SIGMOD Record*, 37(4):115–117, 2008

- [IMP09a] Yannis Ioannidis, Paolo Manghi, and Pasquale Pagano, editors. *Proceedings of the Second Workshop on Very Large Digital Libraries (VLDL 2009), in conjunction with ECDL2008*, Corfu, Greece, October 2009. Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Italy, DELOS Association for Digital Libraries
- [IMP09b] Yannis Ioannidis, Paolo Manghi, and Pasquale Pagano. Second Workshop on Very Large Digital Libraries – VLDL 2009. *D-Lib Magazine*, 15(11/12), November 2009
- [MPI09] Paolo Manghi, Pasquale Pagano, and Yannis Ioannidis. Second workshop on Very Large Digital Libraries – VLDL 2009. *SIGMOD Record*, 38(4):46–48, 2009
- [CMB09] Donatella Castelli, Paolo Manghi, and Stefania Biagioni. OpenAIRE Project - Open Access Infrastructure for Research in Europe, Poster on proceedings of the Eleventh International Conference on Grey Literature (GL11), December 2009
- [KMI⁺10b] Gabriella Kazai, Paolo Manghi, Katerina Iatropoulou, Tim Haughton, Marko Mikulicic, and Natasa Milic-Frayling. ScholarLynk: A Reading List Metaphor for Collaborative Research, Poster on proceedings of the Fifth International Conference of Open Repositories, Madrid, Spain, July 2010
- [KMFH⁺10] Gabriella Kazai, Natasa Milic-Frayling, Tim Haughton, Natalia Manola, Antonis Lempesis, Katerina Iatropoulou, Paolo Manghi, and Marko Mikulicic. Connecting the Local and the Online in Information Management, Poster on proceedings of the 19th International Conference on Information and Knowledge Management, Toronto, Canada, October 2010
- [IMP10] Yannis Ioannidis, Paolo Manghi, and Pasquale Pagano, editors. *Proceedings of the Third Workshop on Very Large Digital Libraries (VLDL 2010), in conjunction with ECDL2010*, Glasgow, Scotland, UK, September 2010. Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Italy, ISTI-CNR
- [CIM11b] Leonardo Candela, Yannis Ioannidis, and Paolo Manghi, editors. *Proceedings of the Fourth Workshop on Very Large Digital Libraries (VLDL 2011), in conjunction with TPDFL2011*, Berlin, Germany, September 2011. Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Italy, ISTI-CNR
- [CIM11a] Leonarco Candela, Yannis Ioannidis, and Paolo Manghi. Fourth workshop on Very Large Digital Libraries. *SIGMOD Record*, 40(4):61–64, 2011

Technical reports and others

- [CCM⁺08] Donatella Castelli, Leonardo Candela, Paolo Manghi, Marko Mikulicic, and Pasquale Pagano. Typed Compound Objects Models for Digital Library Repository Systems. Technical report, Istituto di Scienze e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, October 2008
- [CM10] Donatella Castelli and Paolo Manghi. Open Access Infrastructure for Research in Europe. *Ercim News Magazine - Special Theme: Digital Preservation*, January(80):9–10, 2010

Software Products

- [ACC⁺01] Antonio Albano, Dario Colazzo, Luca Cardelli, Giovanni Conforti, Giorgio Ghelli, Paolo Manghi, and Carlo Sartiani. TQL - Typed Query Language. Product available at: <http://www.di.unipi.it/~ghelli/tql>, Dipartimento di Informatica, Università di Pisa, 2001
- [CGMS06b] Giovanni Conforti, Giorgio Ghelli, Paolo Manghi, and Carlo Sartiani. XPeer - An XML Peer-to-peer database. Product available at: <http://www.di.unipi.it/~manghi/XPeerWeb/homexpeer.htm>, Dipartimento di Informatica, Università di Pisa, 2006
- [MM09] Paolo Manghi and Marko Mikulicic. D-NET Software - DRIVER Network Evolution ToolKit. Product available at: <http://www.d-net.research-infrastructures.eu>, Istituto di Scienze e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, 2009